

# UC Davis

## UC Davis Previously Published Works

### Title

Signatures of Environmental Adaptation During Range Expansion of Wild Common Bean (*Phaseolus vulgaris*)

### Permalink

<https://escholarship.org/uc/item/0m9409c4>

### Authors

Ariani, Andrea  
Gepts, Paul

### Publication Date

2019

### DOI

10.1101/571042

Peer reviewed

1 **Signatures of potential environmental adaptation during**  
2 **range expansion of wild common bean (*Phaseolus vulgaris*)**  
3 **using an integrated landscape genomics approach**

4

5 Andrea Ariani and Paul Gepts\*

6 Department of Plant Sciences, Section of Crop and Ecosystem Sciences,

7 University of California, Davis, CA 95616-8780

8

9 \*Correspondence: [plgepts@ucdavis.edu](mailto:plgepts@ucdavis.edu)

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25 **Abstract**

26 **Background:** Landscape genomics integrates population genetics with  
27 landscape ecology, allowing the identification of the putative molecular  
28 determinants involved in plant environmental adaptation across their natural  
29 geographic and ecological range. Wild *Phaseolus vulgaris*, the progenitor of  
30 common bean (*P. vulgaris*), has a remarkably extended distribution from  
31 northern Mexico to northwestern Argentina. Earlier research has shown that  
32 this distribution represents a range expansion from Mesoamerica to the  
33 southern Andes through several discrete migration events and that the species  
34 colonized areas with different temperature and rainfall compared to its core  
35 area of origin.

36 **Results:** In the current study, we applied a landscape genomics approach to a  
37 collection of 246 wild common bean accessions representative of its broad  
38 geographical and climatic distribution and genotyped for ~20K SNPs. We  
39 applied two different but complementary approaches for identifying loci  
40 putatively involved in environmental adaptation: i) an outlier-detection method  
41 that identifies loci showing strong differentiation between sub-populations; ii)  
42 an association method based on the identification of loci associated with bio-  
43 climatic variables. This integrated approach allowed the identification of  
44 several genes showing signature of selection across the different natural sub-  
45 populations of this species, as well as genes associated with specific bio-  
46 climatic variables related to temperature and precipitation.

47 **Conclusions:** The current study demonstrates the feasibility of landscape  
48 genomics approach for a preliminary identification of novel candidate genes  
49 involved in environmental adaptation in *P. vulgaris*. As a resource for

50 broadening the genetic diversity of the domesticated gene pool of this species,  
51 the genes identified constitute potential molecular markers and introgression  
52 targets for the breeding improvement of domesticated common bean.

53 **Keywords:** Landscape genomics; Crop Wild Relatives (CWRs); climate  
54 adaptation; GWAS; natural selection; domestication

## 55 **Background**

56 Climate change represents one of the primary threats for food security  
57 worldwide, but especially in developing countries that rely heavily on  
58 agricultural production from smallholder farmers [1,2]. Indeed, several studies  
59 have highlighted a predominant role of climate change in reducing agricultural  
60 productivity and increasing inter-annual variability in crop yields, thus directly  
61 affecting food availability and stability [3,4].

62 The increase in average temperatures, along with the higher frequency and  
63 intensity of extreme weather conditions, will require the development of new  
64 plant varieties adapted to this changing environment in order to meet future  
65 food security needs [5,6]. However, the development of new varieties requires  
66 the introduction of genetic diversity into breeding programs to find the correct  
67 combinations of favorable alleles in a specific crop [7]. The genetic variability  
68 available in domesticated plants is generally low due to the bottleneck effect  
69 induced by domestication and subsequent selection during variety  
70 improvement (Zamir 2001; Ford-Lloyd et al. 2011; Gepts 2014), thus new  
71 sources of genetic diversity need to be introduced into breeding programs.  
72 Crop Wild Relatives (CWRs) represent a large, and mostly unexploited, source  
73 of genetic diversity readily available for plant improvement under climate  
74 change [7,8,10,11]. However, the use of CWRs in breeding programs for  
75 improving stress resistance in domesticated species could be hindered by the  
76 lack of knowledge of the genetic determinants of resistance and also by  
77 difficulties in phenotyping a large number of individuals under agricultural  
78 conditions [11–13]. One possible solution for overcoming these difficulties is  
79 the integration of environmental and genotypic datasets to understand the

80 genetic basis of natural selection in wild populations, an approach known as  
81 'landscape genomics' [14,15]. In addition, this approach offers both theoretical  
82 and practical applications since it strengthens the understanding of plant  
83 natural adaptation but allows also the identification of molecular markers that  
84 could be readily applicable for breeding improvement of domesticated plants  
85 [16].

86       Several methods have been developed for identifying signature of natural  
87 selection (e.g., selective sweeps) in natural populations. These methods can be  
88 divided mostly in outlier-detection methods, which identify hard-selection  
89 sweeps, and association methods, that identify soft-selection sweeps [14,17].  
90 Outlier-detection methods are based on population differentiation analysis and  
91 aim at identifying loci with drastic differences in allele frequencies between  
92 populations, as measured by  $F_{st}$  [18,19]. Although based on the assumption  
93 that alleles fixed within sub-populations could confer an evolutionary  
94 advantage in the ecological niche occupied [20,21], these methods do not take  
95 directly into account climatic data and could be biased by complex population  
96 structure and/or demography [22]. On the other hand, association methods  
97 directly correlate genotypic with environmental data and rely on the  
98 assumption that variations of allele frequencies across environmental gradients  
99 are possible signature of local adaptation [23]. The theory beneath  
100 environmental association methods are practically the same as that used in  
101 Genome Wide Association Studies (GWAS) [24] Both approaches employ mixed  
102 model association approaches for correcting for the confounding effects that  
103 could be introduced by population structure and relatedness in the sample  
104 [25].

105 Common bean (*P. vulgaris*) is an essential staple crop providing most of  
106 proteins and micronutrients in the diet of the majority of the population in  
107 several developing countries [26]. The regular consumption of this crop  
108 provides several health benefits, like reducing the risks of heart disease,  
109 obesity, and diabetes [27]. Its cultivation improves agricultural sustainability  
110 thanks to its nitrogen-fixing ability [28]. Common bean shows a surprisingly  
111 high genetic diversity, with the presence of at least three geographically  
112 isolated and divergent wild gene pools located in 1) Mesoamerica and the  
113 northern Andes (MW); 2) the Central Andes (Ecuador and northern Peru; Phl);  
114 and 3) the Southern Andes (southern Peru, Bolivia, and northwestern  
115 Argentina; AW) [29–31]. Common bean was domesticated independently in  
116 Mexico and the Southern Andes, producing locally-adapted varieties and  
117 landraces with specific characteristics [32–38]. The intermediate gene pool in  
118 the Central Andes was not domesticated [39,40]. This wild group has been  
119 recently identified as a cryptic sister species of *P. vulgaris*, named *Phaseolus*  
120 *debouckii*, which was disseminated from the center of origin of this species in  
121 Mesoamerica and remained geographically isolated from the other wild gene  
122 pools of this species [41,42].

123 Wild common bean is an annual vine plants that is found from the state of  
124 Chihuahua in northern Mexico (approx. 35° N. Lat.) to the Córdoba province in  
125 Argentina (approx. 35° S. Lat.), encompassing almost 70 latitudinal degree or  
126 about 10,000 km [43,44]. This species grows in both tropical and sub-tropical  
127 environments across the Americas at elevations between 500 and 2,000 m  
128 a.s.l. with annual rainfall from 500 to 1,800 ml [43,44,12]. This broad  
129 geographic and ecological distribution suggests the existence of genotypes

130 adapted to a wide variety of environmental conditions, which could be useful  
131 donors of abiotic stress resistance for improving domesticated common bean  
132 production under climate change [44,45].

133 Future projection of climate changes under different models predict a  
134 reduction of suitability for common bean production in areas where this plant is  
135 an essential staple crop and also a source of household income, hence  
136 endangering food security and increasing rural poverty in already susceptible  
137 areas of the world [46]. For this reason, it is essential to understand the  
138 molecular mechanisms involved in wild common bean adaptation to different  
139 environments and to identify molecular markers that could be useful in  
140 breeding improvement of this crop. The application of landscape genomics  
141 approaches in wild common bean could help address these issues, as  
142 demonstrated previously in several other plant species like soybean, barley,  
143 *Medicago truncatula*, maize, and *Brachypodium* [16,47–50].

144 In the current study, we applied a landscape genomics approach to  
145 understand environmental adaptation to a dataset comprised of 246 wild  
146 common beans genotyped for ~20K previously developed SNPs [51]. A similar  
147 analysis was performed previously in this species using 148 SNPs located in  
148 genes putatively involved in adaptation to biotic or abiotic stresses [52].  
149 However, the higher number of markers developed in this study and the  
150 broader and more even distribution across the genome of these markers,  
151 results in a more comprehensive and precise analysis of environmental  
152 adaptation in this species. In addition, the genes identified as associated with  
153 environmental variables can be validated and applied in the future for  
154 domesticated common bean breeding improvement.



155

## 156 **Materials and Methods**

### 157 **Plant material and genotype data**

158 A panel of 246 wild *P. vulgaris* accessions, previously genotyped with a  
159 Genotyping-By-Sequencing (GBS) protocol using the *Cvi*AI restriction enzyme  
160 [51], was analyzed. The panel was representative of the ecological and  
161 geographic distribution of this species and included 157 genotypes of the  
162 Mesoamerican (MW), 77 of the Southern Andes (AW), and 12 of the Central  
163 Andes (Northern Peru-Ecuador; PhI) gene pools. The SNPs considered in this  
164 study were those with a Minor Allele Frequency (MAF)  $\geq 0.05$  and less than  
165 20% missing data. The list of the accessions sequenced, with gene pool  
166 information and geographic coordinates, is available in Additional File 1, while  
167 genotyping data in VCF format are available in Additional File 2. The seeds were  
168 provided by the International Center of Tropical Agriculture (CIAT, Cali,  
169 Colombia) and the United States Department of Agriculture Western Regional  
170 Plant Introduction Station (Pullman, WA).

171

### 172 **Spatial analysis**

173 Spatial analyses were conducted within the R statistical environment ([www.r-](http://www.r-project.org)  
174 [project.org](http://www.r-project.org)) using the *dismo* package and its dependencies (*raster* and *sp*). The  
175 geographic coordinates of the individuals analyzed in this study were used for  
176 retrieving the 19 bio-climatic summary variables from the WorldClim database  
177 (<http://www.worldclim.org/>). The data were downloaded at a 30-second  
178 resolution (approximately 0.86 km<sup>2</sup> at the equator). In order to identify a subset  
179 of bio-climatic variables that best summarizes our dataset, we performed a

Principal Component Analysis (PCA) on the scaled and centered variables using the ChemometricsWithR package [53]. We then selected the first two variables with the highest positive and negative loading in the first four principal components (PC1 to PC4) (Additional File 3). Since some of the selected bio-climatic variables showed a high correlation (Additional File 4), we decided to pick only one of the correlated variables for further analysis. The final bio-climatic variables analyzed in this study were: bio\_3 (Isothermality), bio\_5 (Max Temperature of Warmest Month), bio\_6 (Minimum Temperature of Coldest Month), bio\_7 (Temperature Annual Range), bio\_12 (Annual precipitation), bio\_14 (Precipitation of Driest Month), and bio\_18 (Precipitation of Warmest Quarter). In addition to the above-mentioned bio-climatic variables, we included also annual Potential EvapoTranspiration (PET) downloaded from the Global Aridity and PET Database (<http://www.cgiar-csi.org/data/global-aridity-and-pet-database>).

## **Genome scans for selection and association analysis**

Genome scans for selection (i.e., hard selective sweeps) were performed on the final set of SNPs using the pcadapt R package [54], an algorithm able to detect population structure and outlier loci by performing a PCA analysis on SNPgenotypic data. The best number of sub-populations was inferred by visually evaluating the scree plot of eigenvalues for the different principal components (K); the genomic scans for selection were performed for K in the range 2-5. The p-values obtained by this analysis were corrected using the Bonferroni method and only SNPs with a corrected p-value  $\leq 0.001$  were considered as significant.

205 Association analysis (i.e., soft selective sweeps) was performed separately for  
206 each of the seven selected bio-climatic variables and annual PET. For this  
207 analysis, we used the LFMM algorithm [55] implemented in the LEA R package  
208 [56]. This method was developed specifically for identifying signature of  
209 environmental selection in genomic data and is able to efficiently correct for  
210 population history and isolation-by-distance (IBD). In order to correct for  
211 spurious association determined by population structure or IBD, the number of  
212 latent factors (i.e., populations) needs to be decided *a priori* and subsequently  
213 evaluated using the genomic inflation factor parameter. Since LFMM is based  
214 on Monte Carlo Markov Chain (MCMC) sampling, we ran it multiple times for  
215 each association analysis and then averaged the p-values (as suggested in the  
216 software documentation). To identify the best number of populations (K) for  
217 association with each bio-climatic variable, we performed three runs of the  
218 program with K in the range 4-10 and estimated the inflation factor from these  
219 runs [57]. Plots of the inflation factor for different values of K (Additional File 5)  
220 showed that the best inflation factor for reducing False Discovery Rate (FDR)  
221 (i.e., closest to 1) was six for Bio12, Bio14, and Bio5, and 7 for Bio6, Bio18,  
222 Bio7, Bio3, and PET. Based on this preliminary screening, we re-ran the  
223 program with the best number of K for 10 times with 10,000 MCMC iterations  
224 and a burn-in period of 1,000. The p-values were then averaged across the  
225 different runs and corrected using the Bonferroni method. SNPs with a  
226 corrected p-value  $\leq 0.05$  were considered as significant.

227

## 228 **Identification of putatively selected genes**

229 The distance between significant SNPs, identified by genome scans or  
230 association analysis based on the *P. vulgaris* v1.0 genome annotation  
231 (<https://phytozome.jgi.doe.gov/pz/portal.html>), was evaluated using the  
232 GenomicRanges/rtracklayer packages or R [58,59]. Only genes within 5 Kb of a  
233 significant SNPs were chosen as putatively selected genes. This 5 Kb upper  
234 limit was selected based on the genotyping approach used in this study (that  
235 did not allow a full coverage of the genome), but also took into account the  
236 presence of possible regulatory regions immediately adjacent to gene  
237 sequences [60].

238

### 239 **Candidate genes evaluation across genetic groups**

240 For clustering individuals based on genetic groups and visualize allele  
241 frequency variations across clusters, we applied a K-means clustering approach  
242 using the first 5 PCs obtained from pcadapt analysis. As best number of  
243 clusters we selected K=5, as suggested by the scree plot of the eigenvalues  
244 obtained with pcadapt. The clustering analysis was performed using the python  
245 scikit-learn library [61]. For each genetic cluster we calculate allele frequencies  
246 for SNPs tagging candidate genes using VCFtools [62], and plot them on  
247 genetic maps using R.

## 248 **Results**

### 249 **Bio-climatic data analysis**

250 The bio-climatic variables downloaded from the WorldClim database concern  
251 mostly temperature and rainfall during the year. These bio-climatic variables  
252 were developed for generating biologically informative variables useful for  
253 species distribution modeling and landscape genomics approaches. In our  
254 analyses, the 19 bio-climatic variables analyzed showed a great degree of  
255 correlation, in particular for similar variables like bio\_14 (precipitation of the  
256 driest month) and bio\_17 (precipitation of the driest quarter), or bio\_13  
257 (precipitation of the wettest month) with bio\_16 (precipitation of the wettest  
258 quarter) (Additional File 4).

259         The loading plot on the first two PCs showed some correlations between  
260 bio-climatic variables and principal components, as well as strong correlations  
261 between some of the bio-climatic variables analyzed (**Fig. 1A**). In particular,  
262 bio\_12 (annual precipitation) and bio\_4 (temperature seasonality) showed a  
263 strong correlation with PC1. On the other hand, bio\_5 (max temperature of the  
264 warmest month), bio\_8 (mean temperature of the wettest quarter), and bio\_10  
265 (mean temperature of the warmest quarter) showed a strong correlation with  
266 PC2. Interestingly, most of the variables related to precipitation (bio\_12, bio\_14,  
267 bio\_16, bio\_17, bio\_18, and bio\_19) were positively correlated with PC1, the  
268 variables related to seasonal variation (bio\_2, bio\_4, bio\_7, and bio\_15) were  
269 negatively correlated with PC1, while the variables related to temperature  
270 (bio\_1, bio\_5, bio\_8, bio\_9, bio\_10, and bio\_11) were negatively correlated with  
271 PC2.

272 In addition, this PCA on the bio-climatic variables for the genotypes  
273 analyzed showed that the first two principal components (PC1 and PC2)  
274 explained 75% of the variance (**Fig. 1**), while PC1 to PC4 explained  
275 cumulatively > 90% of the variance (Additional File 6). A plot of PC1 vs. PC2  
276 showed some differences in the distribution of the different gene pools of wild  
277 common bean in the PC dimensional space. In particular, the majority of  
278 genotypes from the Mesoamerican (MW1 to MW3) and Intermediate (PhI) gene  
279 pools were distributed towards the positive part of PC1, while the Andean group  
280 were located in the negative part of this axis (**Fig. 1A**). Given the origin of the  
281 genus *Phaseolus* in the Mesoamerican area (with local descendants  
282 represented by MW1 and MW2), three range expansions characterize this  
283 species: 1) PhI, which established wild populations on the western slope of the  
284 Andes in Ecuador and northern Peru; 2) AW, encompassing wild populations in  
285 the southern Andes; and 3) MW3, a more recent and perhaps ongoing  
286 dissemination to Central America and the eastern slope of the northern Andes  
287 [51]. Inspection of Fig. 1A and Additional File 3 shows that the distribution of  
288 the PhI group, which resulted from the earliest range expansion event,  
289 correlates - on bioPC3 - with Isothermality (bio\_3), Temperature Seasonality  
290 (bio\_4), bio\_13 (Precipitation of the Wettest Month), and bio\_18 (Precipitation of  
291 the Warmest Quarter), consistent with a dispersal to an equatorial region. In  
292 contrast, the predominant distribution of the southern Andean accessions (AW)  
293 in the upper left quadrant of Fig. 1 is consistent with earlier observations that  
294 the populations of this gene pool are distributed in cooler and drier locations,  
295 as shown by correlations with bio\_6 (Minimum Temperature of the Coldest  
296 Month), bio\_9 (Mean Temperature of the Driest Quarter), bio\_11 (Mean

297 Temperature of the Coldest Quarter) and bio\_1 (Annual Mean Temperature).  
298 This dissemination occurred with a concomitant lower potential  
299 evapotranspiration [51]. Dispersal of the MW3 group (Fig. 1) increased  
300 Isothermality (bio\_3) and decreased Seasonality (bio\_4) and Precipitation  
301 Seasonality (bio\_15); it also increased Precipitation during the Driest Month  
302 (bio\_14) and Driest Quarter (bio\_17).

303

#### 304 **Genome scan of selection**

305 An analysis of the scree plot of the PCA analysis conducted on SNP data  
306 showed that the majority of the variance could be explained by the first  
307 principal component (15%), even though PC2 to PC5 also explained a  
308 considerable amount of variance in the data (**Fig. 2A**). On the other hand,  
309 after PC5, no significant increase in the explained variance could be detected.  
310 This pattern of the scree plot is representative of a possible range expansion of  
311 this species across the Americas, as hypothesized by a prior evolutionary  
312 analysis of this same collection [51]. Visual inspection of p-value distribution for  
313 genome scans for K=2 and K=3 showed a large proportion of low and high p-  
314 values, while for K=4 and K=5 the distribution of p-values was more uniform,  
315 especially for K=5 (Additional File 7). For this reason, we selected K=5 for  
316 further genome scan analysis.

317 A plot of genetic PCA analysis performed with the pcadapt algorithm was  
318 able to discriminate between the different wild gene pools of this species (**Fig.**  
319 **2B**). In particular, the MW1-MW3 and PhI groups were mostly localized on the  
320 positive part of PC1, while the AW gene pool was localized towards the

321 negative end of this PC. Interestingly, PC1 mostly differentiated MW vs. AW,  
322 while PC2 and PC3 clearly separated the MW+AW groups from the PhI (**Fig. 3**).

323 The genome scan analysis with K=5 identified 84 significant variants  
324 (Bonferroni-corrected p-value  $\leq 0.001$ ) distributed throughout the 11  
325 chromosomes of common bean (**Fig. 4**), tagging 70 annotated genes. The  
326 highest number of tagged genes were identified on chromosomes Pv02 and  
327 Pv04 with 15 and 11 genes, respectively. The genes identified as selected by  
328 genome scan analysis were mostly related to plant development (17 genes),  
329 hormone response (10 genes), ion homeostasis (5 genes), and response to  
330 stress (9 genes) (**Table 1**). Among the genes identified, we found several of  
331 them related to drought and/or abscisic acid (ABA) response like  
332 Phvul.002G331700, a homolog of the *Arabidopsis* KUP6 involved in potassium  
333 uptake transporter and stomata movement; Phvul.002G143100, a glycine-rich  
334 domain protein (GRP) involved in auxin signaling and stress response;  
335 Phvul.004G102800, homologous to *Arabidopsis* SLAH3 involved in ABA  
336 response; Phvul.008G161000, homolog of *Arabidopsis* CAO, a gene related to  
337 chlorophyll biosynthesis and ABA signaling; and Phvul.009G050600, a gene  
338 annotated as an importin  $\beta$  protein involved in ABA and drought response in  
339 *Arabidopsis*.

340



## 341 **Genome-wide association analysis**

342 A genome-wide association analysis identified 49 genes associated with the  
343 bio-climatic variables selected for this analysis. Except for the Bio18  
344 (Precipitation of Warmest Quarter) variable, for which no associations were  
345 detected, the other variables were associated with at least one gene. The bio-  
346 climatic variables with the highest number of associated genes were Bio3  
347 (Isothermality) with 29 genes, and Bio12 (Annual precipitation) with 11 genes  
348 (**Table 2**). The associated genes were located in all 11 common-bean reference  
349 genome chromosomes, except for chromosome Pv06 where there were no  
350 significantly associated SNPs. Interestingly, some of the genes were associated  
351 with more than one bio-climatic variables (**Fig. 5**), suggesting the possibility  
352 that they could be related to multiple environmental stimuli. Among the genes  
353 significantly associated with one or more bio-climatic variables, we found  
354 several of them related to hormone response, ion homeostasis, plant  
355 development, metabolism and response to stress, in particular drought (**Table**  
356 **2**). There was no overlap between the genes identified by genome scan and  
357 association analysis. Among the genes identified, we found some interesting  
358 candidates probably involved in stress resistance, like Phvul.001G034400, a  
359 homolog of *Arabidopsis* KEA6 involved in potassium homeostasis;  
360 Phvul.010G155000, homologous to an *Arabidopsis* phospholipase D  $\alpha$  1  
361 (PLD $\alpha$ 1) involved in ABA signaling; Phvul.010G035200 homolog of a cytokinin  
362 responsive factor homologous of *Arabidopsis*; and Phvul.008G161700,  
363 homologous to an *Arabidopsis* thioredoxin involved in ROS signaling.

364

## 365 **Candidate gene allele distributions**

366 To evaluate the geographic distribution of alleles in candidate genes identified  
367 by genome scan and association analysis, we clustered the genotypes into  
368 groups with a K-means clustering approach on the molecular PCs calculated  
369 with pcadapt. The advantage of a K-means clustering approach, over a  
370 standard population structure analysis, is that it clearly assigns individuals to  
371 specific clusters. The K-means clustering approach identified three clusters for  
372 the MW group, with two clusters (MW1 and MW2) located in Mexico and  
373 another (MW3) in Central America and Colombia, plus one cluster each for the  
374 intermediate (PhI) and the Andean (AW) group (Additional File 8). Interestingly,  
375 the clustering results closely resembled those obtained in a previous study with  
376 more advanced population structure approaches (Additional File 1) [51].

377 The allele frequency distribution of the candidate genes identified by genome  
378 scan showed drastic differentiation between the genetic groups identified (**Fig.**  
379 **6**), as expected from the assumptions of the genome scan approach, with some  
380 alleles being private for just one of the genetic group (like the alternative  
381 alleles for GRP and CAO that are present only in the AW group). On the other  
382 hand, the genes identified by association analysis showed a wide variety of  
383 allele frequencies distribution across the different genetic groups (**Fig. 7**), even  
384 though some genes had only a single allele in some of the populations (like the  
385 reference allele for PLD and TRX in the PhI and AW group). In general, the  
386 genes identified by association analysis showed a higher variation of allele  
387 frequencies in the different MW groups.

388

## 389 **Discussion**

390 Wild common bean (*P. vulgaris*) grows in several areas of Mexico and Central  
391 and South America, from northern Mexico to northwestern Argentina across  
392 ~70 latitudinal degrees, in different environments with a wide range of  
393 altitudes, average temperatures, and rainfall regimes [12,43,44]. Thanks to this  
394 exceptional geographic distribution, its complex evolutionary history, and high  
395 levels of genetic diversity, this species represents an extraordinary resource for  
396 evolutionary studies [29–31,41,42,63,64], but can be also a conceptual  
397 framework for testing and validating landscape genomics approaches in wild  
398 plant populations and its feasibility for breeding improvement of domesticated  
399 crops [16]. In the current study, we identified several genes that could be  
400 involved in environmental adaptation in wild common bean by combining  
401 genome scan and association analysis. If validated, the genes identified could  
402 be useful candidates for improving stress resistance in domesticated common  
403 bean.

404

## 405 **Genome scan of selection**

406 Genetic PCA analysis clearly separated the three groups of this species, as  
407 observed in previous research. In particular, the Intermediate gene pool was  
408 shown again to diverge from the Mesoamerican and Andean gene pools,  
409 especially along the PC2 and PC3, further supporting the hypothesis that this  
410 gene pool is actually a distinct species of *Phaseolus* [41,42]. A genome scan  
411 based on genetic PCA analysis identified several genes with a strong signature  
412 of selection (hard-selection sweep) that could be involved in environmental  
413 adaptation across the geographical range of this species. The identification of

414 several genes involved in plant development and hormone and stress  
415 response, suggests that the different populations of this species adapted to  
416 their environment by integrating and adjusting to developmental, hormonal  
417 and environmental cues. There are several genes among those identified that  
418 could be of interest for improving stress resilience in common bean, like the  
419 KUP6 potassium ( $K^+$ ) transporter located on chromosome Pv02  
420 (Phvul.002G331700). This gene has been directly linked to drought stress by  
421 regulating ABA response and stomata movements in *Arabidopsis* [65]. A  
422 homolog of this gene located on chromosome Pv03 (Phvul.003G052900)  
423 showed a higher genetic and transcriptional diversity in Mesoamerican  
424 domesticated beans than in wild ones [66,67]. Due to the possible role of KUP-  
425 like genes in response to drought stress and their identification as selected  
426 genes in both wild and domesticated populations of common bean, further  
427 studies should focus on the evolution and diversity of this gene family in this  
428 species.

429 Another gene identified in the current study and possibly involved in  
430 adaptation to drought response in wild common bean is Phvul.004G102800,  
431 homolog of SLAH3 of *Arabidopsis*, which was annotated as an S-type anion  
432 channel. This type of channels is rapidly regulated by ABA and stimulate  
433 stomata closure by inhibiting inward  $K^+$  channels, thus reducing  $K^+$  influx into  
434 guard cells [68,69]. In addition to being involved in drought stress response,  
435 this same gene has been recently identified also as related to salinity stress  
436 response in *Arabidopsis* by regulating ion homeostasis between root and shoots  
437 [70].

438       The chlorophyll alpha oxygenase (Phvul.008G161000), identified as a gene  
439   under selection (and a homolog of *Arabidopsis* CAO), has a primary role in the  
440   biosynthesis of chlorophyll b [71]. However, *Arabidopsis* mutants for this gene  
441   showed a reduction of antioxidant compounds (specifically glutathione) in  
442   guard cells and an increased ABA sensitivity in comparison to wild type plants  
443   [72], suggesting a possible involvement of this gene in adaptive response to  
444   stressful environments.

445       Phvul.002G143100, identified as selected within the different sub-  
446   populations of *P. vulgaris*, is annotated as a glycine-rich domain protein,  
447   homologous of *Arabidopsis* GRDP2 gene. GRP are a multi-gene superfamily  
448   present in several organisms, including plants [73]. This gene family has been  
449   associated in plants with several developmental processes and in responses to  
450   both biotic and abiotic stresses [74]. A recent study focusing on the  
451   characterization of the direct *Arabidopsis* homologs of this gene (ATGRDP2)  
452   demonstrated that this gene regulates plant growth and flowering by  
453   accumulating higher level of indole-3-acetic acid and improves abiotic stress  
454   response [75]. In particular, the over-expression of this gene in transgenic  
455   plants increased growth rate and reduced days to flowering. It also increased  
456   salt tolerance in comparison to wild-type plants.

457   In addition to the previous genes identified as selected by genome scan  
458   analysis and putatively involved in environmental response in plants, was  
459   Phvul.009G050600, which is annotated as an importin  $\beta$ -protein homologous to  
460   *Arabidopsis* KPNB1. This gene mediates the import of proteins and protein  
461   complexes between the cytoplasm and the nucleus and is essential in  
462   regulating signal transduction pathways in response to environmental and

463 developmental stimuli [76]. In particular, the *Arabidopsis* homolog of  
464 Phvul.009G050600 (AtKPNB1) has been directly related to ABA and drought  
465 response previously [77].

466

#### 467 **Genome wide association analysis**

468 Association analysis between genotypic data and bio-climatic variables  
469 identified several genes significantly associated with one or more bio-climatic  
470 variables, putatively involved in plant development, ion homeostasis, and  
471 stress response. Among these genes, several could be useful as potential  
472 molecular markers for improving abiotic stress in domesticated common bean.  
473 As examples, we identified a gene related to potassium homeostasis and  
474 annotated as a K<sup>+</sup> efflux antiporter (KEA) gene associated with Bio12 (Annual  
475 Precipitation) and Bio7 (Temperature Annual Range). Potassium is an essential  
476 macronutrient involved in several physiological and developmental processes  
477 in all living organism, and in plants this cation is also essential in maintaining  
478 plant osmotic potential, cytosolic pH, and stomata movement [78,79]. In  
479 addition, variation in K<sup>+</sup> homeostasis is one of the first response to several  
480 abiotic and biotic stresses in plants, allowing the plants to rapidly respond to  
481 stressful conditions [80] and making the KEA gene identified in the current  
482 study an interesting candidate gene for further analysis.

483 Another gene, significantly associated with Bio14 (Precipitation of Driest Month)  
484 is Phvul.010G155000, which is annotated as a phospholipase (PLD $\alpha$ 1). These  
485 genes are involved in the biosynthesis of phosphatidic acid (PA), which is an  
486 important signaling molecule in response to several stresses in plants [81]. In  
487 particular, PA is involved in the ABA signaling cascade and regulates stomata

closure in plants by directly interacting and blocking ABI1, an inhibitor of ABA response in plants [82]. This gene regulates stomatal closure and ABA-dependent hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) production in *Vicia faba* as well [83], making this gene an interesting candidate for improving drought response in common bean.

An additional gene, significantly associated with Bio5 (Max Temperature of Warmest Month), is Phvul.010G035200, annotated as a cytokinin response factor homologous of *Arabidopsis* CRF4. Cytokinin is an essential plant hormone involved in growth and developmental processes [84,85], but in recent years it has also been implicated in the response and adaptation to different environmental stresses [86,87]. CRF genes are a class of plant transcription factors responsive to cytokinin that integrate hormonal and environmental signals for adapting plant growth and development in response to the environment [88,89]. The *Arabidopsis* homolog of this gene has been previously related to acclimation to cold temperatures [90]. Since this gene has been associated with temperature variables in wild common bean, it could also be involved in adaptation to temperature variation in this species.

Another gene of interest, Phvul.008G161700, is significantly associated with Bio3 (Isothermality) and is annotated as a thioredoxin protein. These proteins are involved in the regulation of oxidative stress response and in scavenging reactive oxygen species (ROS) in plants [91]. Other than being simple byproducts of cellular metabolism, ROS molecules have been recognized as important signaling molecules that regulate the response to several environmental stresses in plants [92,93]. Due to their ability to control the redox state of the cell, thioredoxin represents a key component of the ROS

513 signal transduction pathways in plants and in the response to environmental  
514 stress [94]. Thus, this gene could constitute another interesting candidate gene  
515 for improving stress resistance in domesticated common bean.

## 516 517 **Comparison of genes identified by genome scan and GWAS**

518 Even though the genes identified by outlier-detection methods (hard-selection  
519 sweeps) and association methods (soft-selection sweeps) are involved in  
520 similar processes, there is no overlap between the candidate genes identified  
521 by the two approaches. This could be the direct results of the different  
522 assumptions underlying these methods. Indeed, genome scan analysis identify  
523 genes that shows drastic variations of allele frequencies between natural  
524 subpopulations [14,17]. This approach is independent from bio-climatic  
525 variables, thus the SNPs identified as under selection by this analysis could be  
526 the results of selective mechanisms not taken into account by association  
527 analysis, like soil composition, pathogen pressure and/or competition with other  
528 plants. On the other hand, association analyses identify SNPs showing slight  
529 variations in allele frequencies across environmental gradients that can  
530 increase environmental adaptation in natural populations [14,17]. This  
531 selection process usually acts on natural standing variations and favor the  
532 presence of multiple alleles and haplotypes, instead of allele fixation within  
533 populations [95].

## 534 535 **Conclusions**

536 In conclusion, landscape genomic analysis of wild common bean genotypes  
537 allowed us to identify several genes showing a signature of presumed selection



538 in this species. It is likely that two methods – genome scan and GWAS - are  
539 indeed complementary for understanding local adaptation in wild plant  
540 populations, as observed previously in other species [49,96] and are a feasible  
541 approach for the preliminary identification of novel candidate genes for  
542 adaptation to climatic differences along the exceptionally broad habitat of wild  
543 common bean. Further corroboration of the actual role of the candidate genes  
544 in adaptation will come from introgression of these genes from wild to  
545 domesticated beans and a concurrent phenotypic analysis showing improved  
546 performance under stress conditions.

547

#### 548 **Abbreviations**

549 CWRs: Crop Wild Relatives; GWAS: Genome Wide Association Studies; SNPs:  
550 Single Nucleotide polymorphisms; GBS: Genotyping-By-Sequencing; MW:  
551 Mesoamerican wild gene pool; AW: Andean wild gene pool; PhI: Northern Peru-  
552 Ecuador gene pool; PCA: Principal Component Analysis; PET: Potential  
553 EvapoTranspiration; LFMM: Latent Factor Mixed Models; IBD: isolation-by-  
554 distance; MCMC: Monte Carlo Markov Chain; ABA: abscisic acid; GRP: glycine-  
555 rich domain protein; PLD $\alpha$ 1: phospholipase D  $\alpha$  1; PA: phosphatidic acid; ROS:  
556 reactive oxygen species

557

#### 558 **Acknowledgements**

559 This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC  
560 Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and  
561 S10RR027303. We thank the Genetic Resources Unit of the Centro Internacional  
562 de Agricultura Tropical (CIAT, Cali, Colombia) and the Western Regional Plant

563 Introduction Station of the USDA (Pullman, WA) for providing samples of wild *P.*  
564 *vulgaris* used in this study.

565

## 566 **Funding**

567 This project was supported by Agriculture and Food Research Initiative (AFRI)  
568 Competitive Grant No. 2013-67013-21224 from the USDA National Institute of  
569 Food and Agriculture.

570

## 571 **Availability of Data and Materials**

572 Raw sequencing data are available at the NCBI Sequence Read Archive ([http://](http://www.ncbi.nlm.nih.gov/sra)  
573 [www.ncbi.nlm.nih.gov/sra](http://www.ncbi.nlm.nih.gov/sra)) under the accession numbers SRX2771627 and  
574 SRX2771628. The variants file and the relative geographical coordinates used  
575 for performing the analysis are available as additional files of the current  
576 manuscript.

577

## 578 **Authors' contributions**

579 AA performed the experiment, analyzed the data and wrote the manuscript. PG  
580 designed the experiment, supervised the work and wrote the manuscript. All  
581 authors approved the final version of the manuscript.

582

## 583 **Ethics approval and consent to participate**

584 Not applicable

585

## 586 **Consent for publication**

587 Not applicable

588

589 **Competing interests**

590 The authors declare that they have no competing interests.

591 **References**

- 592 1. Rippke U, Ramirez-Villegas J, Jarvis A, Vermeulen SJ, Parker L, Mer F, et al.  
593 Timescales of transformational climate change adaptation in sub-Saharan  
594 African agriculture. *Nat Clim Change*. 2016;6:605–9.
- 595 2. Campbell BM, Vermeulen SJ, Aggarwal PK, Corner-Dolloff C, Girvetz E,  
596 Loboguerrero AM, et al. Reducing risks to food security from climate change.  
597 *Glob Food Secur*. 2016;11:34–43.
- 598 3. Wheeler T, Braun J von. Climate change impacts on global food security.  
599 *Science*. 2013;341:508–13.
- 600 4. Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N. A meta-  
601 analysis of crop yield under climate change and adaptation. *Nat Clim Change*.  
602 2014;4:287–91.
- 603 5. Lobell DB, Burke MB, Tebaldi C, Mastrandrea MD, Falcon WP, Naylor RL.  
604 Prioritizing climate change adaptation needs for food security in 2030. *Science*.  
605 2008;319:607–10.
- 606 6. IPCC. Managing the risks of extreme events and disasters to advance climate  
607 change adaptation: special report of the Intergovernmental Panel on Climate  
608 Change. Cambridge University Press; 2012.
- 609 7. Ford-Lloyd BV, Schmidt M, Armstrong SJ, Barazani O, Engels J, Hadas R, et al.  
610 Crop wild relatives—undervalued, underutilized and under threat? *BioScience*.  
611 2011;61:559–65.
- 612 8. Zamir D. Improving plant breeding with exotic genetic libraries. *Nat Rev*

613 Genet. 2001;2:983–9.

614 9. Gepts P. The contribution of genetic and genomic approaches to plant  
615 domestication studies. Curr Opin Plant Biol. 2014;18:51–9.

616 10. Spillane C, Gepts P. Evolutionary and genetic perspectives on the dynamics  
617 of crop genepools. (2001); Available from:  
618 <http://agris.fao.org/agris-search/search.do?recordID=XF2003411459>

619 11. Brozynska M, Furtado A, Henry RJ. Genomics of crop wild relatives:  
620 expanding the gene pool for crop improvement. Plant Biotechnol J.  
621 2016;14:1070–85.

622 12. Cortés AJ, Monserrate FA, Ramírez-Villegas J, Madriñán S, Blair MW. Drought  
623 tolerance in wild plant populations: the case of common beans (*Phaseolus*  
624 *vulgaris* L.). PLoS ONE. 2013;8:e62898.

625 13. Zhang H, Mittal N, Leamy LJ, Barazani O, Song B. Back into the wild—Apply  
626 untapped genetic diversity of wild relatives for crop improvement. Evol Appl.  
627 2016;10:5–24.

628 14. Schoville SD, Bonin A, François O, Lobreaux S, Melodelima C, Manel S.  
629 Adaptive genetic variation on the landscape: methods and cases. Annu Rev  
630 Ecol Evol Syst. 2012;43:23–43.

631 15. Bragg JG, Supple MA, Andrew RL, Borevitz JO. Genomic variation across  
632 landscapes: insights and applications. New Phytol. 2015;207:953–67.

633 16. Anderson JE, Kono TJY, Stupar RM, Kantar MB, Morrell PL. Environmental  
634 association analyses identify candidates for abiotic stress tolerance in *Glycine*

- 635 *soja*, the wild progenitor of cultivated soybeans. *G3 GenesGenomesGenetics*.  
636 2016;6:835–43.
- 637 17. Wagner HH, Fortin M-J. A conceptual framework for the spatial analysis of  
638 landscape genetic data. *Conserv Genet*. 2012;14:253–61.
- 639 18. Wright S. The genetical structure of populations. *Ann Eugen*. 1949;15:323–  
640 54.
- 641 19. Lewontin RC, Krakauer J. Distribution of gene frequency as a test of the  
642 theory of the selective neutrality of polymorphisms. *Genetics*. 1973;74:175–95.
- 643 20. Haldane JBS. A mathematical theory of natural and artificial selection. (Part  
644 VI, Isolation.). *Math Proc Camb Philos Soc*. 1930;26:220–30.
- 645 21. Kimura M. On the probability of fixation of mutant genes in a population.  
646 *Genetics*. 1962;47:713–9.
- 647 22. Narum SR, Hess JE. Comparison of  $F_{ST}$  outlier tests for SNP loci under  
648 selection. *Mol Ecol Resour*. 2011;11 Suppl 1:184–94.
- 649 23. Manel S, Joost S, Epperson BK, Holderegger R, Storfer A, Rosenberg MS, et  
650 al. Perspectives on the use of landscape genetics to detect genetic adaptive  
651 variation in the field. *Mol Ecol*. 2010;19:3760–72.
- 652 24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common  
653 diseases and complex traits. *Nat Rev Genet*. 2005;6:95–108.
- 654 25. Lipka AE, Kandianis CB, Hudson ME, Yu J, Drnevich J, Bradbury PJ, et al. From  
655 association to prediction: statistical methods for the dissection and selection of

656 complex traits in plants. *Curr Opin Plant Biol.* 2015;24:110–8.

657 26. Gepts P, Aragão FJL, Barros E de, Blair MW, Brondani R, Broughton W, et al.  
658 genomics of *Phaseolus* beans, a major source of dietary protein and  
659 micronutrients in the Tropics. In: Moore PH, Ming R, editors. *Genomics Trop Crop*  
660 *Plants*. Springer New York; 2008 p. 113–43.

661 27. Messina V. Nutritional and health benefits of dried beans. *Am J Clin Nutr.*  
662 2014;100 Suppl 1:437S–42S.

663 28. Rubiales D, Mikic A. Introduction: legumes in sustainable agriculture. *Crit*  
664 *Rev Plant Sci.* 2015;34:2–3.

665 29. Chacón SMI, Pickersgill B, Debouck DG, Arias JS. Phylogeographic analysis  
666 of the chloroplast DNA variation in wild common bean (*Phaseolus vulgaris* L.) in  
667 the Americas. *Plant Syst Evol.* 2007;266:175–95.

668 30. Koenig R, Gepts P. Allozyme diversity in wild *Phaseolus vulgaris*: further  
669 evidence for two major centers of genetic diversity. *Theor Appl Genet.*  
670 1989;78:809–17.

671 31. Mamidi S, Rossi M, Moghaddam SM, Annam D, Lee R, Papa R, et al.  
672 Demographic factors shaped diversity in the two gene pools of wild common  
673 bean *Phaseolus vulgaris* L. *Heredity.* 2013;110:267–76.

674 32. Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, et al.  
675 Molecular analysis of the parallel domestication of the common bean  
676 (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol.*  
677 2013;197:300–13.

- 678 33. Blair MW, Soler A, Cortés AJ. Diversification and population structure in  
679 common beans (*Phaseolus vulgaris* L.). PLOS ONE. 2012;7:e49488.
- 680 34. Chacón SMI, Pickersgill B, Debouck DG. Domestication patterns in common  
681 bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean  
682 cultivated races. Theor Appl Genet. 2005;110:432–44.
- 683 35. Gepts P, Osborn TC, Rashka K, Bliss FA. Phaseolin-protein variability in wild  
684 forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for  
685 multiple centers of domestication. Econ Bot. 1986;40:451–68.
- 686 36. Mamidi S, Rossi M, Annam D, Moghaddam S, Lee R, Papa R, et al.  
687 Investigation of the domestication of common bean (*Phaseolus vulgaris*) using  
688 multilocus sequence data. Funct Plant Biol. 2011;38:953–67.
- 689 37. Rossi M, Bitocchi E, Bellucci E, Nanni L, Rau D, Attene G, et al. Linkage  
690 disequilibrium and population structure in wild and domesticated populations of  
691 *Phaseolus vulgaris* L. Evol Appl. 2009;2:504–22.
- 692 38. Singh SP, Gepts P, Debouck DG. Races of common bean (*Phaseolus*  
693 *vulgaris*, Fabaceae). Econ Bot. 1991;45:379–96.
- 694 39. Debouck DG, Toro O, Paredes OM, Johnson WC, Gepts P. Genetic diversity  
695 and ecological distribution of *Phaseolus vulgaris* (fabaceae) in Northwestern  
696 South America. Econ Bot. 1993;47:408–23.
- 697 40. Kami J, Velásquez VB, Debouck DG, Gepts P. Identification of presumed  
698 ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc Natl Acad Sci.  
699 1995;92:1101–4.



- 700 41. Rendón-Anaya M, Herrera-Estrella A, Gepts P, Delgado-Salinas A. A new  
701 species of *Phaseolus* (Leguminosae, Papilionoideae) sister to *Phaseolus*  
702 *vulgaris*, the common bean. *Phytotaxa*. 2017;313:259–66.
- 703 42. Rendón-Anaya M, Montero-Vargas JM, Saburido-Álvarez S, Vlasova A,  
704 Capella-Gutierrez S, Ordaz-Ortiz JJ, et al. Genomic history of the origin and  
705 domestication of common bean unveils its closest sister species. *Genome Biol*.  
706 2017;18:60.
- 707 43. Gepts P. Origin and evolution of common bean: past events and recent  
708 trends. *HortScience*. 1998;33:1124–30.
- 709 44. Porch TG, Beaver JS, Debouck DG, Jackson SA, Kelly JD, Dempewolf H. Use  
710 of wild relatives and closely related species to adapt common bean to climate  
711 change. *Agronomy*. 2013;3:433–61.
- 712 45. Acosta-Gallegos JA, Kelly JD, Gepts P. Prebreeding in common bean and use  
713 of genetic diversity from wild germplasm. *Crop Sci*. 2007;47:S-44.
- 714 46. Ramirez-Cabral NYZ, Kumar L, Taylor S. Crop niche modeling projects major  
715 shifts in common bean growing areas. *Agric For Meteorol*. 2016;218–219:102–  
716 13.
- 717 47. Westengen OT, Berg PR, Kent MP, Brysting AK. Spatial structure and  
718 climatic adaptation in African maize revealed by surveying SNP diversity in  
719 relation to global breeding and landrace panels. *PloS One* 2012;7:e47832.
- 720 48. Yoder JB, Stanton-Geddes J, Zhou P, Briskine R, Young ND, Tiffin P. Genomic  
721 signature of adaptation to climate in *Medicago truncatula*. *Genetics*

2014;196:1263–75.

49. Dell’Acqua M, Zuccolo A, Tuna M, Gianfranceschi L, Pè ME. Targeting environmental adaptation in the monocot model *Brachypodium distachyon*: a multi-faceted approach. *BMC Genomics* 2014;15:801.

50. Abebe TD, Naz AA, Léon J. Landscape genomics reveal signatures of local adaptation in barley (*Hordeum vulgare* L.). *Front Plant Sci.* 2015;6.

51. Ariani A, Berny Mier Y Teran J, Gepts P. Spatial and temporal scales of range expansion in wild *Phaseolus vulgaris*. *Mol Biol Evol.* 2017;35:119-131.

52. Rodriguez M, Rau D, Bitocchi E, Bellucci E, Biagetti E, Carboni A, et al. Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol.* 2016;209:1781–94.

53. Wehrens R. *Chemometrics with R*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011

54. Luu K, Bazin E, Blum MGB. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *bioRxiv.* 2016;056135.

55. Frichot E, Schoville SD, Bouchard G, François O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol.* 2013;30:1687–99.

56. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol.* 2015;6:925–9.

- 743 57. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*.  
744 1999;55:997–1004.
- 745 58. Lawrence M, Gentleman R, Carey V. rtracklayer: an R package for  
746 interfacing with genome browsers. *Bioinformatics*. 2009;25:1841–2.
- 747 59. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al.  
748 Software for computing and annotating genomic ranges. *PLoS Comput Biol*.  
749 2013;9:e1003118.
- 750 60. Li X, Zhu C, Yeh C-T, Wu W, Takacs EM, Petsch KA, et al. Genic and nongenic  
751 contributions to natural variation of quantitative traits in maize. *Genome Res*.  
752 2012;22:2436–44.
- 753 61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al.  
754 Scikit-learn: machine learning in Python. *J Mach Learn Res*.  
755 2011;12:2825–2830.
- 756 62. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The  
757 variant call format and VCFtools. *Bioinforma Oxf Engl*. 2011;27:2156–8.
- 758 63. Bitocchi E, Nanni L, Bellucci E, Rossi M, Giardini A, Zeuli PS, et al.  
759 Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by  
760 sequence data. *Proc Natl Acad Sci*. 2012;109:E788–96.
- 761 64. Kwak M, Gepts P. Structure of genetic diversity in the two major gene pools  
762 of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor Appl Genet*.  
763 2009;118:979–92.
- 764 65. Osakabe Y, Arinaga N, Umezawa T, Katsura S, Nagamachi K, Tanaka H, et

765 al. Osmotic stress responses and plant growth controlled by potassium  
 766 transporters in Arabidopsis. Plant Cell 2013;25:609–24.

767 66. Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al.  
 768 Decreased nucleotide and expression diversity and modified coexpression  
 769 patterns characterize domestication in the common bean. Plant Cell Online.  
 770 2014;26:1901–12.

771 67. Bitocchi E, Rau D, Bellucci E, Rodriguez M, Murgia ML, Gioia T, et al. Beans  
 772 (*Phaseolus* ssp.) as a model for understanding crop evolution. Front Plant Sci.  
 773 2017;8.

774 68. Geiger D, Maierhofer T, AL-Rasheid KAS, Scherzer S, Mumm P, Liese A, et al.  
 775 Stomatal closure by fast abscisic acid signaling is mediated by the guard cell  
 776 anion channel SLAH3 and the receptor RCAR1. Sci Signal. 2011;4:ra32-ra32.

777 69. Zhang A, Ren H-M, Tan Y-Q, Qi G-N, Yao F-Y, Wu G-L, et al. S-type anion  
 778 channels SLAC1 and SLAH3 Function as essential negative regulators of inward  
 779 k<sup>+</sup> channels and stomatal opening in Arabidopsis. Plant Cell Online  
 780 2016;28:949–65.

781 70. Cubero-Font P, Maierhofer T, Jaslan J, Rosales MA, Espartero J, Díaz-Rueda P,  
 782 et al. Silent S-type anion channel subunit SLAH1 gates SLAH3 open for chloride  
 783 root-to-shoot translocation. Curr Biol. 2016;26:2213–20.

784 71. Espineda CE, Linford AS, Devine D, Brusslan JA. The AtCAO gene, encoding  
 785 chlorophyll a oxygenase, is required for chlorophyll b synthesis in *Arabidopsis*  
 786 *thaliana*. Proc Natl Acad Sci. 1999;96:10507–11.

787 72. Jahan MS, Nozulaidi M, Khairi M, Mat N. Light-harvesting complexes in  
788 photosystem II regulate glutathione-induced sensitivity of Arabidopsis guard  
789 cells to abscisic acid. *J Plant Physiol.* 2016;195:1-8.

790 73. Sachetto-Martins G, Franco LO, de Oliveira DE. Plant glycine-rich proteins: a  
791 family or just proteins with a common motif? *Biochim Biophys Acta*  
792 2000;1492:1-14.

793 74. Mangeon A, Junqueira RM, Sachetto-Martins G. Functional diversity of the  
794 plant glycine-rich proteins superfamily. *Plant Signal Behav.* 2010;5:99-104.

795 75. Ortega-Amaro MA, Rodríguez-Hernández AA, Rodríguez-Kessler M,  
796 Hernández-Lucero E, Rosales-Mendoza S, Ibáñez-Salazar A, et al.  
797 Overexpression of AtGRDP2, a novel glycine-rich domain protein, accelerates  
798 plant growth and improves stress tolerance. *Front Plant Sci.* 2015;5.

799 76. Merkle T. Nucleo-cytoplasmic partitioning of proteins in plants: implications  
800 for the regulation of environmental and developmental signalling. *Curr Genet.*  
801 2003;44:231-60.

802 77. Luo Y, Wang Z, Ji H, Fang H, Wang S, Tian L, et al. An Arabidopsis homolog  
803 of importin  $\beta$ 1 is required for ABA response and drought tolerance. *Plant J.*  
804 2013;75:377-89.

805 78. Shabala S. Regulation of Potassium Transport in Leaves: from Molecular to  
806 Tissue Level. *Ann Bot.* 2003;92:627-34.

807 79. Sharma T, Dreyer I, Riedelsberger J. The role of K<sup>+</sup> channels in uptake and  
808 redistribution of potassium in the model plant *Arabidopsis thaliana*. *Front Plant*

809 Sci. 2013;4.

810 80. Shabala S, Pottosin I. Regulation of potassium transport in plants under  
811 hostile conditions: implications for abiotic and biotic stress tolerance. *Physiol*  
812 *Plant*. 2014;151:257–79.

813 81. Saucedo-García M, Gavilanes-Ruíz M, Arce-Cervantes O. Long-chain bases,  
814 phosphatidic acid, MAPKs, and reactive oxygen species as nodal signal  
815 transducers in stress responses in *Arabidopsis*. *Front Plant Sci*. 2015;6.

816 82. Zhang W, Qin C, Zhao J, Wang X. Phospholipase D $\alpha$ 1-derived phosphatidic  
817 acid interacts with ABL1 phosphatase 2C and regulates abscisic acid signaling.  
818 *Proc Natl Acad Sci* 2004;101:9508–13.

819 83. Qu Y, An Z, Zhuang B, Jing W, Zhang Q, Zhang W. Copper amine oxidase  
820 and phospholipase D act independently in abscisic acid (ABA)-induced stomatal  
821 closure in *Vicia faba* and *Arabidopsis*. *J Plant Res*. 2014;127:533–44.

822 84. Durán-Medina Y, Díaz-Ramírez D, Marsch-Martínez N. Cytokinins on the  
823 Move. *Front Plant Sci*. 2017;8.

824 85. Kieber JJ. Cytokinins. *Arab Book Am Soc Plant Biol*. 2002;1.

825 86. Novakova M, Dobrev P, Motyka V, Gaudinova A, Malbeck J, Pospisilova J, et  
826 al. Cytokinin function in drought stress response and subsequent recovery.  
827 *Biotechnol Sustain Agric 2006 Beyond*. Springer, Dordrecht; 2007 p. 171–4.

828 87. O'Brien JA, Benková E. Cytokinin cross-talking during biotic and abiotic  
829 stress responses. *Front Plant Sci*. 2013;4.

- 830 88. Kim J. CYTOKININ RESPONSE FACTORs gating environmental signals and  
831 hormones. *Trends Plant Sci.* 2016;21:993–6.
- 832 89. Rashotte AM, Goertzen LR. The CRF domain defines Cytokinin Response  
833 Factor proteins in plants. *BMC Plant Biol.* 2010;10:74.
- 834 90. Zwack PJ, Compton MA, Adams CI, Rashotte AM. Cytokinin response factor 4  
835 (CRF4) is induced by cold and involved in freezing tolerance. *Plant Cell Rep.*  
836 2016;35:573–84.
- 837 91. Gelhaye E, Rouhier N, Navrot N, Jacquot JP. The plant thioredoxin system.  
838 *Cell Mol Life Sci* 2005;62:24–35.
- 839 92. D’Autréaux B, Toledano MB. ROS as signalling molecules: mechanisms that  
840 generate specificity in ROS homeostasis. *Nat Rev Mol Cell Biol.* 2007;8:813–24.
- 841 93. Sewelam N, Kazan K, Schenk PM. Global plant stress signaling: reactive  
842 oxygen species at the cross-road. *Front Plant Sci.* 2016;7.
- 843 94. Sevilla F, Camejo D, Ortiz-Espín A, Calderón A, Lázaro JJ, Jiménez A. The  
844 thioredoxin/peroxiredoxin/sulfiredoxin system: current overview on its redox  
845 function in plants and regulation by reactive oxygen and nitrogen species. *J Exp*  
846 *Bot.* 2015;66:2945–55.
- 847 95. Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of  
848 adaptation from standing genetic variation. *Genetics* 2005;169:2335–52.
- 849 96. Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J. Complex patterns of  
850 local adaptation in teosinte. *Genome Biol Evol.* 2013;5:1594–609.

851 97. Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, et al. A  
852 reference genome for common bean and genome-wide analysis of dual  
853 domestications. Nat Genet. 2014;46:707-13.



854 **Figure legends**

855 **Figure 1** Bio-climatic data analysis. **(A)** Loading plot of the PCA analysis. **(B)**  
856 Principal Component Analysis (PCA) of the bio-climatic data. Groups are colored  
857 according to the K-mean clustering analysis conducted in this study, which  
858 gave results very similar to the STRUCTURE analysis conducted by Ariani et al.  
859 (2016): MW1, MW2, and MW3: Mesoamerican wild gene pools; AW: Andean wild  
860 gene pool; PhI: Intermediate wild gene pool.

861 **Figure 2** PCA analysis on molecular data. **(A)** Screeplot of the PCA explained  
862 variance. **(B)** PCA plot based on molecular data of the different genotyped  
863 analyzed in the current study. MW: Mesoamerican Wild; AW: Andean Wild; PhI:  
864 Intermediate gene pool.

865 **Figure 3** Three-dimensional plot of the PCA analysis on molecular data. Points  
866 are colored as in Figure 2A.

867 **Figure 4** Manhattan plot of the genome scan data with 5 sub-populations (K).  
868 The blue dashed line represents the significance threshold (Bonferroni p-value  
869  $\leq 0.001$ ).

870 **Figure 5** Chromosome ideogram of the genes identified as associated with the  
871 bio-climatic variables analyzed. Only chromosomes with significantly  
872 associated variants are shown. Each circle represents a different bio-climatic  
873 variable. When available, gene annotations are shown. The centromeric regions  
874 shown are based on the results from [97].

875 **Figure 6** Allele frequency distribution across different genetic groups for  
876 candidate genes identified by genome scan analysis. *P. vulgaris* v1.0 genes  
877 annotation and ID: **(A)** Potassium uptake transporter (Phvul.002G331700); **(B)**  
878 Glycine-rich domain protein (Phvul.002G143100); **(C)** ABA response

879 (Phvul.004G102800); **(D)** Chlorophyll biosynthesis and ABA signaling  
880 (Phvul.008G161000); **(E)** ABA and drought response (Phvul.009G050600). For  
881 panel **(E)** the PhI group was removed due to complete missing data on the  
882 SNPs. REF: Reference allele, ALT: Alternative allele, according to the *P. vulgaris*  
883 v1.0 gene version.

884 **Figure 7** Allele frequency distribution across different genetic groups for  
885 candidate genes identified by association analysis. *P. vulgaris* v1.0 genes  
886 annotation and ID: **(A)** Potassium efflux antiporter (Phvul.001G034400); **(B)**  
887 Phospholipase D  $\alpha$  1 (Phvul.010G155000); **(C)** Cytokinin responsive factor  
888 (Phvul.010G035200); **(D)** Thioredoxin (Phvul.008G161700). Reference and  
889 Alternative alleles are colored as in Figure 6.

890

#### 891 **Additional Files**

892 **Additional File 1 (Additional\_File\_1.csv):** List of the *P. vulgaris* accessions  
893 sequenced in the current study, with gene pool information, geographic  
894 coordinates, gene pools and genetic group information.

895 **Additional File 2 (Additional\_File\_2.vcf):** Genotypic data of the *P. vulgaris*  
896 accessions analyzed in the current study

897 **Additional File 3 (Additional\_File\_3.csv):** Eigenvalues of the different  
898 bioclimatic variables along the first four principal components.

899 **Additional File 4 (Additional\_File\_4.pdf):** Correlation graphs between bio-  
900 climatic variables for the different *P. vulgaris* accessions analyzed. Correlation  
901 coefficients are rendered using circles (upper-right part) or by showing directly  
902 the value (lower-left part). Color are based on color-bar in the right side of the  
903 graph.

904 **Additional File 5 (Additional\_File\_5.pdf):** Plots of the inflation factor for  
905 different values of K across the climatic variables selected for association study  
906 **Additional File 6 (Additional\_File\_6.pdf):** Cumulative variance explained  
907 by the different PCs when performing a PCA on bio-climatic variables  
908 **Additional File 7 (Additional\_File\_7.pdf):** P-values distribution for genome  
909 scans with 2 (A), 3 (B), 4 (C) or 5 (D) sub-populations.  
910 **Additional File 8 (Additional\_File\_8.pdf):** Plot of geographic distribution of  
911 the wild *P. vulgaris* analyzed in the current studies. Genotypes are colored  
912 based on the different clusters identified by K-means clustering (Additional File  
913 1).